Journal of Intelligent Fuzzy Systems 1 (2024) 1–15
IOS Press

Examining the Influence of Socio-economic Factors on Male and Female Students in Mathematic Performance

Ben Johnson
Technological University Dublin, Ireland
E-mail: X00229603@mytudublin.ie

Abstract. This study investigates the influence of various day to day amenities & parental occupation and education, on mathematics performance for secondary and university education in engineering students, using data from the Colombian Institute for the Evaluation of Education (ICFES). By analyzing patterns in students' score, we aim to determine how various attributes, such as access to facilities, parental education, and critical reading score, impact maths results. Statistical and machine learning methods were employed to categorize and quantify these effects, offering information on the design of more effective educational assessments. We created a variety of models with relatively good accuracy, the Random Forest model had an accuracy of 71%, while Logistic Regression achieved 68%. Our findings provide valuable information for educators, curriculum developers, and educational tool designers, enabling them to refine question structure to enhance student engagement and comprehension in mathematics.

Keywords: Mathematics, Education, Machine learning, Statistical analysis, Ethical AI

1. Introduction

The performance of students in mathematics, a foundational subject with significant applications across multiple academic and professional disciplines, is influenced by a broad range of factors, including social, economic, and household characteristics. Research indicates that engagement in assessments, particularly performance-based tasks, can positively impact learning outcomes. When students actively participate in assessments that allow them to demonstrate problem-solving or project-based skills, they gain ownership of their learning, enhancing motivation and performance. These factors, combined with broader social influences, can foster a positive feedback loop between engagement and achievement in mathematics [1].

This study investigates how social and economic variables—such as parental education, household income, and access to technology—affect mathematics performance among higher education engineering students in general, but also in male and female subgroups. Understanding these relationships is critical for improving assessment practices and ensuring equitable evaluation across diverse student populations.

The research uses a national assessment dataset to explore how demographic and socioeconomic characteristics shape students' math performance. Key factors such as social class, household revenue in terms of monthly minimum wage, and internet access are examined to identify patterns that influence academic outcomes. Given the importance of math skills for engineering students' academic success and future careers, this study highlights potential disparities stemming from varied social backgrounds.

By linking socioeconomic variables to mathematics performance, the study contributes to a broader understanding of how assessment designs can be more inclusive. Insights gained from this analysis can guide the development of fairer evaluation methods that reflect students' diverse experiences and learning en-

vironments. The research aims to inform educational policies and practices that support student success regardless of their social and economic circumstances.

Guided by the frameworks outlined in Creswell & Creswell's Research Design [34], this quantitative study will use data analysis techniques to examine the impact of key social features on students' mathematics scores. The dataset includes many valuable features with a particular focus on demographic subgroups. The research excludes unrelated features to maintain a targeted analysis of socioeconomic influences.

This study's methodology will involve data processing, feature encoding, and the application of classification models such as Random Forest and Logistic Regression to identify the most significant predictors of mathematics performance. By analyzing these variables, the study aims to provide executable recommendations that can guide the design of assessments and educational strategies in higher education for both female and male students equally. Ultimately, the goal is to offer a deeper understanding of how various social and economic factors influence academic performance and to propose ways in which assessment practices can be adapted to support diverse student populations more effectively.

2. Literature Review

Mathematics education has long been recognized as a fundamental component of academic development, particularly in higher education, where it shapes critical skills such as problem-solving, logical reasoning, and analytical thinking. However, significant disparities in mathematics performance persist across various educational systems, institutions, and student demographics. This section reviews the existing body of research on mathematics education, with a particular focus on patterns in student performance, the impact of socio-demographic and socioeconomic variables, and the methodologies employed to analyze these relationships, such as those used in the current study with a national assessment dataset.

In recent years, the field of education has been dramatically transformed, particularly with the onset of the COVID-19 pandemic. The rapid shift to online learning, or "e-learning," has altered the educational landscape by expanding access to learning materials and platforms. This shift has also led to an increase in the availability of student data, similar to the data used in this study, which can be analyzed to identify patterns in student performance across different socio-demographic groups. Online learning platforms, such as Chegg and Brightspace, which cater to millions of students worldwide, have provided valuable insights into how students engage with mathematics content, highlighting the role that digital access plays in shaping performance. While these platforms can be beneficial, it is important to recognize that not all students have equal access to technology, and this disparity can be an additional barrier to success for students from lower socio-economic backgrounds. [15,16,17]

Despite the potential benefits of e-learning, online education also comes with its challenges, particularly in terms of student performance. Research has shown that factors such as socio-economic background, access to facilities, and household circumstances, such as the number of available cars or internet access, can all impact student performance. The dataset that we used includes variables related to household items and internet access, which will allow for an in-depth examination of how these factors influence mathematics performance. However, it is also important to acknowledge that the data collected may not fully capture all of the complex features influencing performance, such as student gender, learning disabilities, or emotional factors like stress. Future research should explore how these variables interact with socio-demographic characteristics to provide a more comprehensive understanding of the factors that affect math performance [18,19].

The emotional and psychological aspects of learning also play a critical role in mathematics performance. Anxiety, both related to mathematics and to test-taking in general, has been shown to significantly hinder academic achievement. One meta-analysis found that test anxiety, particularly in mathematics, has a moderating effect on performance, with students reporting higher levels of stress and lower performance as the difficulty of math problems increases. This finding underscores the importance of considering emotional factors when designing assessments and evaluating student performance. While the data the we used does not include direct measures of anxiety or emotional states, it would be beneficial in future work to examine how these psychological variables, along with socio-demographic factors, contribute to performance disparities in mathematics. By analyzing how stress, anxiety, and other emotions interact with socio-economic

background and educational access, researchers can gain a more nuanced understanding of the factors influencing math performance and assessment outcomes [20].

One study using data from the Educational Longitudinal Study of 2002–2006 found that, within the current research sphere, there exists an assumption that all students whose data were used in the study had received similar benefits from taking advanced math courses, irrespective of their socioeconomic status or racial/ethnic backgrounds. The study revealed that the net effect of taking advanced math courses was more beneficial for students with low socioeconomic status (SES) compared to their higher SES counterparts. This finding challenges the assumption of uniform benefits from advanced math education and highlights the potential disparities in the impact of such courses. The dataset utilized in our study, which includes variables such as what jobs the student's parents hold, their highest level of education and overall household income, can be leveraged to investigate how these factors contribute to the benefits of advanced mathematics education. By examining these relationships, we can better understand how students from various, diverse backgrounds experience and benefit from advanced math courses, which is crucial for improving educational practices and outcomes in mathematics education [14].

Information and Communication Technology (ICT) has become a pervasive element in students' daily lives across Europe, with 96% of young Europeans using the internet regularly. Previous research has shown that the impact of ICT on mathematics and science performance varies depending on student attitudes, confidence, and the autonomous use of ICT. While some studies have identified negative correlations between ICT use and academic performance, others have found that a higher level of student engagement with ICT—particularly when students feel confident and perceive the technology as useful—correlates with improved math and science performance. Notably, the study also found that while a country's GDP had little effect on student performance, the availability of extracurricular activities did have a positive influence. The dataset in this study provides a valuable opportunity to examine how variables like access to the internet, household resources, and other socio-economic factors contribute to students' engagement with ICT and, consequently, their mathematics performance. By exploring the intersection of ICT access and social factors, we can gain insights into how digital access influences student learning and performance in mathematics [21,22].

In a 2020 study, researchers examined the predictive effects of mathematics proficiency compared to English language proficiency in predicting university performance. The study concluded that mathematics skills had a stronger relationship with university success than verbal skills, with higher mathematical proficiency being particularly predictive of achieving a first-class honours degree and excelling in STEM courses. This aligns with the importance of mathematics in higher education, especially for students pursuing STEM fields. The ICFES dataset, which includes social features and keywords related to mathematical & language competency, offers a unique opportunity to investigate the specific areas of everyday life where students from different backgrounds struggle the most. A selection of ICFES data is shown in Figure 1 as a correlation matrix, where we can clearly see that language competency plays a larger part in maths performance than other features. By analyzing these features and more, this research can provide valuable insights into the specific mathematical concepts that students find challenging and allow for more targeted interventions to improve performance in these areas. The focus on mathematics proficiency as a key indicator of academic success underscores the importance of understanding how socio-demographic and educational factors—such as parental education and revenue—affect students' mathematical abilities and their potential for achieving high academic standards in higher education [8].

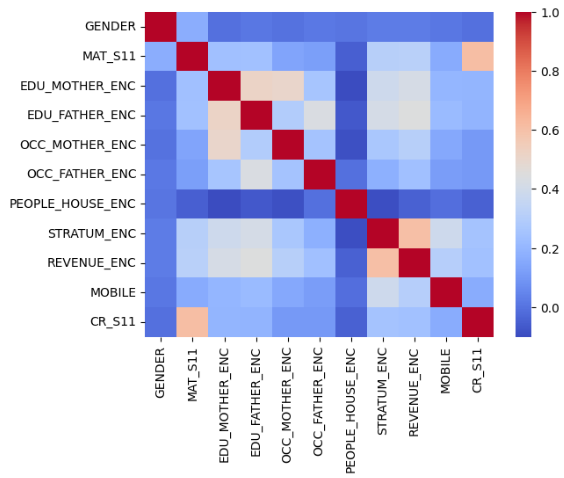

Fig. 1. Data Selection - Correlation Matrix

Furthermore, many countries have exhibited significant variation in mathematics performance, as demonstrated by large-scale assessments such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS). These differences are influenced by a range of systemic factors, including curriculum design, teacher quality, and assessment practices. Monseur and Lafontaine [3] highlight how variations in educational systems can lead to disparities in student outcomes, particularly in mathematics. Similarly, research by Schukajlow et al. [4] has emphasized the importance of task authenticity in fostering student engagement and improving mathematical problem-solving performance. This is particularly relevant when considering the features available in this study's dataset, such as the familiarity with various problem types or the alignment between real-world contexts and the content being taught. The inclusion of socio-demographic variables, like parental education and income, in the dataset allows for a closer examination of how these systemic factors might affect mathematics performance across different student populations. Understanding the role of task authenticity, along with the socio-economic and demographic influences captured in the data, can offer deeper insights into how contextual factors shape student performance in mathematics.

Specific attributes of mathematics questions, such as difficulty level, topic, and linguistic complexity, have been shown to have a substantial impact on student performance. Heinze and Reiss [5] demonstrated that a student's success in solving mathematics problems is closely linked to their familiarity with question formats and their ability to decode the language of instruction. This is particularly relevant to the current study, which explores how students' demographic characteristics and social backgrounds influence their ability to perform in mathematics. Research by Prediger et al. [6] examined the alignment between curriculum content and assessment frameworks and how this alignment affects the fairness and validity of mathematics tests. The dataset used in this study includes variables such as parental education, household income, and other socio-economic indicators, which can be analyzed to investigate how these factors influence students' understanding and performance in mathematics, particularly when they are faced with complex or unfamiliar question formats. By examining how these socio-demographic factors interact with task attributes such as difficulty and language, the study can shed light on the challenges different groups of students face when engaging with mathematics assessments.

European studies have also highlighted the significant role of cultural and regional factors in shaping educational outcomes, particularly in mathematics. Research by Rindermann and Baumeister [35] underscores the disparities in mathematical competence between Western and Eastern European countries, attributing these differences to variations in educational policies, resources, and cultural attitudes towards education. Additionally, the impact of socioeconomic status and language background on student performance has been extensively studied. Findings indicate that these variables significantly mediate students' ability to perform well in standardized mathematics assessments (Stankov et al. [7]). The data used within this study captures some of these socio-economic and demographic factors, such as mother and father education levels and household income, which provide an opportunity to explore how such disparities manifest in student performance. By linking these variables to student outcomes in mathematics, this research aims to contribute to the growing body of work on the impact of socio-economic background and regional differences on educational achievement. Investigating these factors will allow for a more nuanced understanding of how socioeconomic and cultural elements contribute to the disparities in mathematics performance observed across Europe.

Various statistical and machine learning techniques have been widely employed to analyze mathematics performance, offering valuable insights into the predictors of student success and areas for targeted intervention. Traditional methods, such as regression analysis and equation modeling, have long been used to identify key factors influencing student performance, including socio-demographic variables and academic background (Schukajlow et al. [4]). More recently, advanced approaches like data mining and machine learning techniques have gained prominence for their ability to uncover hidden patterns in large datasets, which may not be immediately apparent through traditional methods. Prediger et al. [6] discuss how these techniques allow researchers to explore complex interactions between question features, student demographics, and contextual variables. This is particularly relevant to the dataset used in this study, which includes socio-demographic factors such as parental education, household income, and access to resources.

By applying machine learning models (e.g., Random Forest, Gradient Boost) to this dataset, this research aims to identify how these socio-economic and contextual variables interact with the achievement, scale, and other characteristics of math results, thereby offering a more nuanced understanding of the factors that influence student performance.

Despite the extensive body of research on mathematics education, few studies have systematically examined the influence of socioeconomic factors on performance using large-scale datasets such as the Colombian Institute for the Evaluation of Education dataset. Additionally, there is a notable gap in the literature regarding cross-class analyses that account for regional differences in educational practices and question design, or the comparison between male and female students. This study aims to address these gaps by employing both statistical and machine learning methods to analyze how various social attributes—such as income, social class, and the household amenities—impact student performance in higher education settings. By utilizing the ICFES dataset, which includes key socio-demographic and contextual variables, this research seeks to provide practical findings into how different student groups perform on mathematics assessments. These findings will be valuable to educators, curriculum developers, and assessment designers, offering evidence-based recommendations for creating more equitable and effective assessment practices across different educational contexts.

3. Methodology

3.1. Data Exploration

The dataset used in this study was provided by the Colombian Institute for the Evaluation of Education and sourced from a comprehensive national assessment study [2]. It comprises a significant number of instances and variables, making it an ideal candidate for AI-based modeling and analysis. The dataset includes 12,411 rows and 19 columns, with the primary target variable being "MAT_S11," which represents the math score for each student. This dataset offers a broad spectrum of data from students across different socioeconomic backgrounds, critical reading performance, and demographic categories. The target variable, math performance (MAT_S11), is continuous and varies between 0 and 100, providing a solid basis for regression-based modeling and further analysis. The charts shown in Figure 2 represent the distribution of math scores, overall, in terms of Legal Minimum Monthly Wage (LMMW, in the dataset), and kernel density estimation by social class, or Stratum, per student. Also shown is a heatmap of math score against critical reading score. As mentioned in section 2, there is scope to analyze the relationship between language comprehension and mathematic performance.

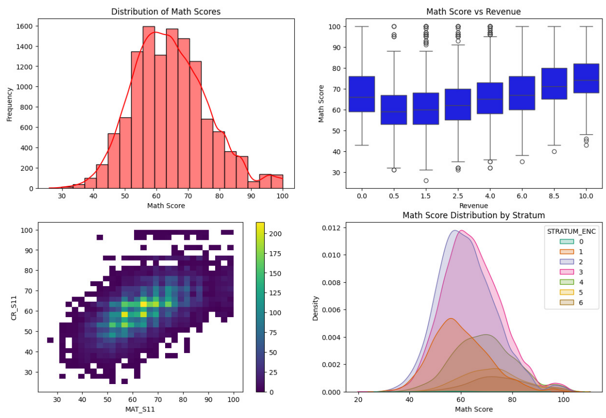

Fig. 2. Exploratory Data Analysis

While the target variable is continuous, other variables in the dataset display some imbalances. The "gender" variable, for example, is slightly skewed towards male students, who represent 59% of the data. However, the education, occupation and revenue variables exhibit some skewness, with a higher proportion of students coming from families with medium to high education levels and income brackets. This distribution is important to consider as it may influence the ability to generalize findings across all socioeconomic classes.

Upon inspecting the dataset, it was observed that certain categorical columns, including the identification and SISBEN columns would not be required for the scope of this project. For instance, the SISBEN variable classifies students according to a vulnerability assessment. We decided that these columns can be dropped on the basis of already having enough information to create a suitable model. We will retain the most informative variables, categorical and numerical, for further analysis.

Additionally, categorical variables such as "gender," "people_house," parental education columns and

household items & services will require encoding for machine learning purposes. Label encoding or one-hot encoding will be applied to transform these categorical features into numerical formats that can be used effectively in machine learning models. The encoding process is crucial for handling the large set of categorical variables present in the dataset, allowing us to better explore relationships between demographic, socioeconomic factors, and student performance in mathematics.

3.2. Data Preprocessing

In preparation for modelling, we had to employ some data cleaning and manipulation steps to the issues identified above. The dataset contains 0 null values, 0 duplicates and was generally quite clean. The column names had to be normalized, for example, during processing, we stripped and capitalized the columns, and added _ENC to the end in order to signify encoding. This maintained readability within the code and helped to avoid any future issues while training the model that could have risen due to the spaces contained within the column names.

Additionally, we decided to drop 22 columns that were related to SISBEN, Quantitative Reasoning, Formulation of Engineering Projects and other unrelated features. These columns are valuable as administrative columns, but they are not suitable or statistically significant, for the scope of this research.

Most of the columns provided valuable, socioeconomic data about the students, including the mathematic score achieved. The preprocessing steps involved:

1. Data inspection for nulls & duplicates
2. Label-encoding the parental education data, as it is ordinal, we wanted to maintain the natural order of educational achievement by each parent.
3. Parental occupation data: We will do one run with one-hot encoded variables, then try label encoding to see if there is any difference or improvement.
4. Binary encoding the household services columns, the data was presented as Yes or No, so we simply mapped 1 or 0 to the value.
5. Label encoding the rest of the columns like revenue & stratum.
6. Creating visualizations & brief statistical analysis

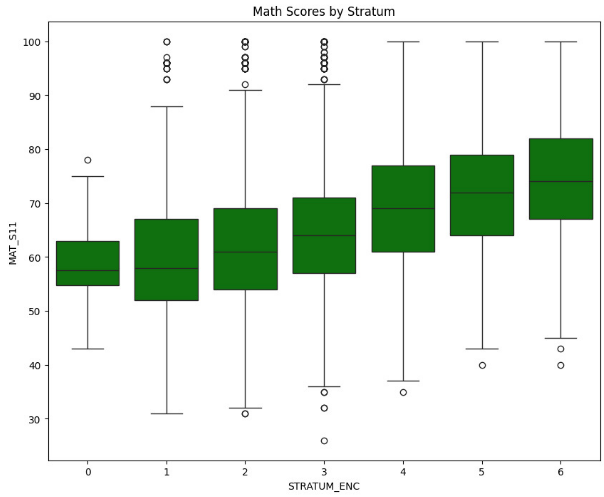

Fig. 3. Maths score against Stratum class

Colombia's stratum system (sistema de estratificación) is a unique socioeconomic classification method used to categorize residential areas and determine utility pricing and public service costs. The stratum system divides residential neighborhoods into six different socioeconomic levels, ranging from Stratum 1 (lowest) to Stratum 6 (highest) [9]. This classification is based on the physical characteristics of the housing and the surrounding neighborhood, including:

- Building materials
- Urban infrastructure
- Access to public services
- Neighborhood amenities
- Environmental characteristics

This feature could be key in this research, as seen in Figure 3 above. There is a clear upward trend in mathematic performance depending on which Stratum a student is classified in.

3.3. Model Options

- Linear Regression: Linear regression creates a relationship between the dependent variable (Y) and one or more independent variables (X) (also known as regression line) using the best fit straight line [10].

- Logistic Regression (LR): Logistic regression uses the logistic (sigmoid) function to estimate probabilities and works well with linearly separable datasets. It can overfit high-dimensional data,

but regularization techniques can help prevent this. A major limitation is its assumption of linearity between dependent and independent variables. While logistic regression can be used for both classification and regression, it is more commonly applied to classification problems. [10]

- Decision Tree (DT): Decision Trees are simple algorithms for classification tasks. Instances are classified by evaluating the attribute defined at each node, starting from the root node and following the branch that corresponds to the attribute value. The most commonly used criteria for splitting are "gini" for Gini impurity and "entropy" for information gain. DT are easy to understand and can handle both categorical and numerical data. However, they can be prone to overfitting if not properly tuned. [10,11]
- Random Forest Classifier (RFC): Random Forest uses a method called "parallel ensembling" to combine multiple decision trees on different dataset sub-samples to improve performance and minimize overfitting. RF can be used for classification or regression problems, as it adapts well to both categorical and continuous variables. [10]
- GradientBoost (GB): GB models are similar to RFC, int he sense that the use a learning algorithm that generates a final model based on previous models. The gradient is used to minimize the loss function. [10]

The choice of algorithm for regression and classification problems depends on various factors, including dataset size, complexity, and desired outcome. Linear Regression is a fundamental technique for modeling linear relationships between variables, while Logistic Regression is well-suited for binary classification tasks. Decision Trees and Random Forests offer flexibility in handling both categorical and numerical data, and are effective in capturing complex patterns. Gradient Boosting provides a powerful framework for building accurate predictive models by iteratively improving upon previous models. The optimal algorithm for a specific problem usually requires experimentation and careful consideration of its strengths and limitations.

3.4. Model Development

The model development process followed a systematic pipeline, including data preprocessing, feature engineering, model training, and evaluation. We used several machine learning algorithms to predict students' math scores, leveraging regression and classification techniques. Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), and classification accuracy guided model performance assessment. We cleaned and prepared the dataset by:

- Handling Missing Values: Replacing missing values in parental education, occupation, and household features with appropriate categories or zeros.
- Encoding Categorical Variables: Ordinal encoding was applied to educational levels and household income. Binary features such as TV, Internet, and car ownership were encoded as 0/1 indicators.
- Standardization: Numerical features were standardized where necessary.

3.4.1. Regression Models

- Linear Regression: Baseline model trained on all features. Metrics: MSE, RMSE, MAE, and R^2. Feature importance was determined from absolute coefficient values, shown in Figure 4.

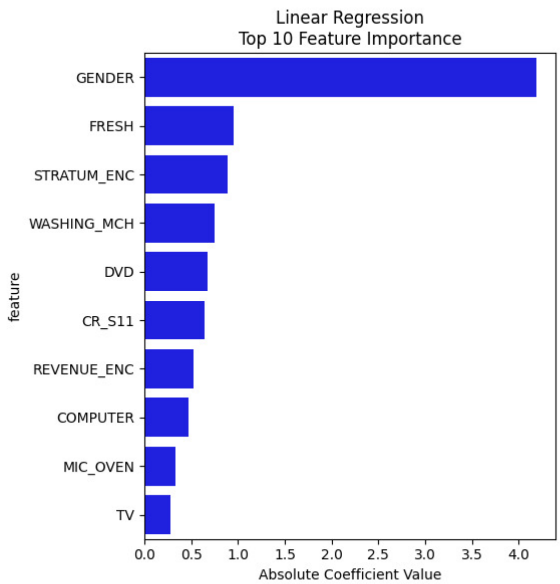

Fig. 4. LinearRegression Feature Importance

- Polynomial Regression: Model extended with polynomial features (degree 2). Improved performance at the cost of increased complexity.

– Gradient Boosting Regressor: Model trained with default hyperparameters. Feature importance identified critical factors such as parental education levels and household income.

– Hyperparameter Tuning: Performed RandomizedSearchCV to optimize Gradient Boosting. Tuned parameters: learning rate, max depth, and number of estimators.

– XGBoost Regressor: Similar configuration to Gradient Boosting, offering increased flexibility and better performance through advanced boosting.

3.4.2. Classification Models

To explore different perspectives, math scores were categorized into three classes: Low, Medium, and High. We tested two classifiers:

– Logistic Regression: One-vs-Rest (OvR) approach for multiclass classification. Performance evaluated using classification reports and confusion matrices.

– Random Forest Classifier: Built an ensemble model using 100 decision trees. Evaluated using accuracy, precision, recall, and F1-scores.

3.5. Evaluation Metrics

– Accuracy: Accuracy is amount of correct calls (true-positive and true-negative) that were made in proportion to total dataset. Higher results generally indicate a better overall predictive model, but it is highly situational and needs to be evaluated on a case by case basis. [12]

– Precision: Expresses the percentage of correctly classified instances in the set of all the instances, the truly positive instances, the truly negative instances, or the instances classified as positive, respectively. [13]

– Recall: True-positive rate is also known as recall or sensitivity and is the probability of the model detecting a truly positive case. False-negative rate is proportion of positive cases missed. A high recall value indicates less false negatives. [12]

– F1-Score: The function of the true-positive rate and positive predictive value (precision and recall) to give an overall indication of performance of the classifier. [12]

– Mean Squared Error (MSE): The MSE is used to calculate the amount of variance between true values and predicted values. The MSE close the line of best fit is to the set of points [30]

– Mean Absolute Error (MAE): he average of the absolute difference between the observed value and the predicted values. The difference between the MAE and MSE is that MAE takes the absolute difference between the predicted values and the observed values whereas the MSE takes the squared difference. [30]

– Root Mean Squared Error (RMSE): The square root of the mean of squares of all the errors. In other words, the Standard deviation of the errors. [30]

4. Results

The first model applied to predict student mathematics scores was a Linear Regression model, with One-hot encoded parental occupation values. The dataset included 19 predictor variables related to students' demographic, socioeconomic, and household characteristics. After training and evaluating the model, the following performance metrics were obtained:

Metric	**Value**
Mean Squared Error (MSE)	77.6833
Root Mean Squared Error (RMSE)	8.8138
Mean Absolute Error (MAE)	6.8805
R-squared Score (R^2)	0.4355

Table 1
LinearRegression Performance Metrics

The RMSE of 8.8138 indicates that, on average, the predicted math scores deviate from the actual scores by approximately 8.82 points on a 0-100 scale. The MAE of 6.89 further confirms that typical prediction errors are in this range. The R^2 score of 0.4355 suggests that the model explains approximately 43.5% of the variance in students' math scores based on the provided features. While this indicates moderate predictive power, it also highlights that a significant portion of the variability remains unexplained by the linear model, suggesting potential non-linear relationships in the dataset.

Next, we tried to transform the original features into polynomial features of degree 2. This enables

the model to fit non-linear relationships between features and the target variable. The results were not optimal when using One-hot encoding, of which results suggested serious issues during model training. We switched back to Label encoding the parental occupation, and ran the PolynomialFeatures model again, producing similar results to standard Linear Regression:

Metric	Value
Mean Squared Error (MSE)	78.3451
Root Mean Squared Error (RMSE)	8.8513
R-squared Score (R^2)	0.4368

Table 2

LinearRegression (Polynomial=2) Performance Metrics

The Mean Squared Error of 78.35 and Root Mean Squared Error of 8.85 suggest that the model's predictions deviate from actual math scores by approximately 8.85 points on average. The R-squared value of 0.44 indicates that the model explains 44% of the variance in the target variable, leaving 56% unexplained. While the model captures some patterns, there is room for improvement, possibly through more complex models, feature engineering, or hyperparameter tuning.

The next model we used was a GradientBoostRegressor, which is an iterative collection of sequentially arranged tree models so as the next model learns from the error of the former model [31].

Metric	Value
Mean Squared Error (MSE)	76.9723
Root Mean Squared Error (RMSE)	8.7734
Mean Absolute Error (MAE)	6.8584
R-squared Score (R^2)	0.4467

Table 3

GradientBoostRegressor Model Performance Metrics

The lower MSE, RMSE, and MAE values suggest better prediction accuracy compared to the Linear Regression model. A higher R^2 score of 0.4467 means the Gradient Boosting model explains about 44.7% of the variance in math scores, outperforming Linear Regression (43.4%). We also had a look at the feature importance to determine which features have the most predictive power, shown in Figure 5.

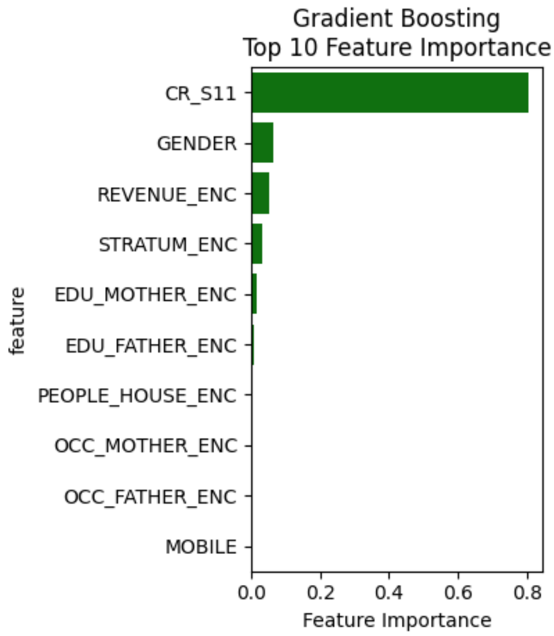

Fig. 5. GradientBoosting Feature Importance

In an attempt to improve the GradientBoost model performance, we tuned some hyperparamters and utilized RandomizedSearchCV. This strategy takes as input a machine learning model, a distribution of hyperparameters, and a cross-validation strategy. The distribution of hyperparameters specifies how to sample values from each hyperparameter range [32]. Results are shown in Table 4.

Metric	Value
Mean Squared Error (MSE)	76.8690
Root Mean Squared Error (RMSE)	8.7675
Mean Absolute Error (MAE)	6.8604
R-squared Score (R^2)	0.4474

Table 4

GradientBoost (RandomizedSearchCV) Performance Metrics

The Gradient Boosting Regressor with RandomizedSearchCV tuning shows improved predictive performance compared to the polynomial regression model. The MSE of 76.87 and RMSE of 8.77 indicate slightly better accuracy, reducing prediction errors by about 0.08 points. The MAE of 6.86 highlights that typical prediction errors are smaller than the RMSE, suggesting limited impact from extreme outliers. The R-squared score of 0.45 shows that the model explains 45% of the variance in math scores, reflecting a marginal improvement in explanatory power.

The final regression model improvement that we tried was using Extreme Gradient Boosting (XGBoost). It is a scalable and optimized algorithm in computer science that improves the speed and prediction performance of Gradient Boosting Machines (GBM) [33]. The XGBoost Regression results are similar to the Linear Regression model and slightly worse than Gradient Boosting. All three models results are shown in Table 5.

Metric	LinearRegression	GradientBoosting	XGBoost
MSE	77.88	76.97	79.04
RMSE	8.83	8.77	8.89
MAE	6.89	6.86	6.95
R^2	0.4340	0.4467	0.4319

Table 5
Model Performance Comparison for Math Score Prediction

To further explore the predictive capabilities of machine learning models, we categorized students' math scores into three distinct bins. These bins were chosen according to the existing distribution of maths results in the dataset, with a distinct bell-curve shape peaking around the mean value, 64.32.

- Low: Scores below 55
- Medium: Scores between 55 and 75
- High: Scores of 75 and above

The classification models tested were Logistic Regression and Random Forest Classifier, while implementing a StandardScaler. These models aimed to predict the categorical math score labels using the 19 selected features. The StandardScaler standardizes the features by removing the mean and scaling to unit variance. This step ensures that features with larger numerical ranges do not disproportionately influence the model, which is crucial for the performance of models like Logistic Regression that rely on the magnitude of coefficients.

Class	Precision	Recall	F1-Score	Support
High	0.70	0.71	0.27	430
Low	0.69	0.24	0.35	526
Medium	0.66	0.94	0.77	1527
Accuracy	0.66 (2483)			
Macro avg	0.68	0.45	0.47	2483
Weighted avg	0.67	0.66	0.60	2483

Table 6
Classification Report for LogisticRegression

The Logistic Regression model achieved an overall accuracy of 66%, indicating moderate performance. The model demonstrates strong precision in identifying high-scoring students (0.70), suggesting that when it predicts a high score, it is often correct. However, it struggles with recall in this category (0.17), meaning it misses many actual high-scoring students. The model shows similar performance in the low-scoring category, with a precision of 0.69 but a low recall of 0.24. This suggests that while it is accurate in its low-score predictions, it often fails to identify true low-scoring students. The model performs best in the medium-scoring category, with a high recall of 0.94 and a reasonable precision of 0.66. This indicates that it effectively identifies medium-scoring students but may misclassify some as high or low. While the macro average F1-score of 0.47 suggests a balanced performance across categories, the weighted average F1-score of 0.60, influenced by the majority class, highlights the model's overall tendency to favor the majority class.

The Random Forest Classifier achieved an overall accuracy of 64.36%, which is slightly lower than the Logistic Regression model. However, it exhibits a more balanced performance across the three classes.

- High-Scoring Students: The model has a precision of 0.54, meaning that when it predicts a high score, it's correct 54% of the time. The recall of 0.28 indicates that it only identifies 28% of the actual high-scoring students.
- Low-Scoring Students: The model has a precision of 0.57 and a recall of 0.37. While it's relatively accurate in its low-score predictions, it still misses a significant number of actual low-scoring students.
- Medium-Scoring Students: The model performs the best in this category, with a high recall of 0.84 and a precision of 0.67. This suggests that it effectively identifies medium-scoring students but may misclassify some as high or low.

Overall, the Random Forest model, while showing improvement in balanced performance, still faces challenges in accurately identifying high-scoring students.

For female students, the model achieves an accuracy of 64%, slightly lower than for males. Similar to the male results, the "High" category performs well, with a precision of 0.73 and recall of 0.81, yielding an F1-score of 0.77. However, the recall for the "Low" category is again low at 0.17, resulting in an F1-score of 0.25. The "Medium" category performs

relatively better, with an F1-score of 0.50. The overall trend shows that while the model can predict high-performing students reasonably well, it struggles with low-performing students, which may be attributed to class imbalance or insufficient data for the lower categories.

Class	Precision	Recall	F1-Score	Support
High	0.49	0.16	0.24	135
Low	0.65	0.47	0.54	247
Medium	0.69	0.87	0.77	627
Accuracy	0.68 (1009)			
Macro avg	0.61	0.50	0.52	1009
Weighted avg	0.65	0.68	0.64	1009

Table 7
Classification Report for Female students (RFC)

The Random Forest Classifier, when applied to female students, exhibited varying performance across different binning strategies. With the initial binning of 40, 40-59, and 60+, the model demonstrated a reasonable ability to predict high-performing students, as evidenced by the high precision and recall for the "High" category. However, it struggled with low-performing students, particularly in terms of recall, potentially due to class imbalance or insufficient data.

When the binning was adjusted to 55, 55-75, and 75+, the model's performance improved slightly. While the precision remained relatively consistent, the recall increased for all categories, especially for the "Medium" category. This suggests that the revised binning better aligned with the distribution of female students' math scores, leading to more accurate predictions.

Comparing the performance of the model for female and male students, both groups exhibited similar trends. Both models struggled to accurately identify low-performing students, indicating a potential limitation in the model's ability to capture subtle differences in performance at the lower end of the spectrum. However, both models performed relatively well in predicting high-performing students.

Overall, the Random Forest Classifier proved to be a useful tool for predicting student performance, but its accuracy was influenced by factors such as binning strategy and potential class imbalance. Future research could explore more advanced techniques, such as feature engineering or ensemble methods, to further improve the model's predictive capabilities.

Class	Precision	Recall	F1-Score	Support
High	0.56	0.37	0.44	303
Low	0.57	0.30	0.39	259
Medium	0.68	0.84	0.75	912
Accuracy	0.65 (1474)			
Macro avg	0.60	0.50	0.53	1474
Weighted avg	0.63	0.65	0.62	1474

Table 8
Classification Report for Male Students (RFC)

5. Conclusion

5.1. Discussion

The use of Artificial Intelligence in educational research, especially when dealing with sensitive data like the one used in this study, raises important ethical considerations. As AI systems increasingly assist in data analysis, ensuring that the models used are transparent, fair, and free from bias is crucial. This is particularly pertinent when studying student performance across various demographic groups, as the risk of inadvertently reinforcing existing inequalities through biased algorithms is significant. Research has shown that machine learning models are susceptible to biases present in the training data, which could lead to skewed results that do not fairly represent all student groups [23].

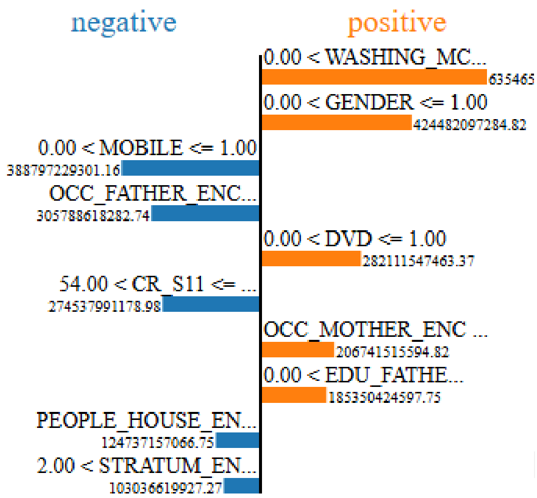

Fig. 6. LIME Results Sample (id=276)

In the context of this study, factors such as occupation and education levels, household income, and access to resources are critical predictors of math per-

formance, as identified in Figure 6, showing the Local Interpretable Model-agnostic Explanations (LIME) results of a single instance. In this isolated case, access to amenities like a washing machine, mobile phone connection and DVD player had increased influence on the final result. If these socio-demographic variables are not carefully considered during data preprocessing or model training, the AI system may generate results that unfairly disadvantage students from lower socio-economic backgrounds. Therefore, it is essential to apply ethical AI practices, such as fairness-aware algorithms and regular audits for model bias, to ensure that the results generated do not perpetuate or exacerbate social inequalities [24]. Moreover, as AI continues to play an increasing role in educational decision-making, frameworks such as the EU's "Ethics Guidelines for Trustworthy AI" must be considered to ensure the responsible deployment of AI in educational research.

As AI applications grow in education, issues surrounding data privacy and security become increasingly important. The General Data Protection Regulation (GDPR) in the European Union provides a robust framework for the collection and use of personal data, including educational datasets like the one in this study. GDPR mandates that personal data must be processed lawfully, transparently, and for specific purposes [25].

The dataset used in this study, which includes sensitive socio-economic information, requires rigorous safeguards to ensure that student privacy is respected and that the data is used only for the intended research purposes. Ethical AI deployment involves compliance with data protection laws to protect individuals from potential harm arising from data misuse or exposure. In the context of AI in education, this also means ensuring that the AI models do not inadvertently expose sensitive student information through predictive analytics or data mining [26]. Therefore, any AI-driven study in education must implement data anonymization and secure data handling practices to uphold the privacy and rights of students, aligning with both ethical guidelines and legal requirements.

5.2. Interpretation of Results & Future Work

The interpretation of results in studies involving AI and machine learning in education requires careful consideration of the context, limitations, and potential biases embedded within the data. In this study, which examines the impact of socio-economic and demographic factors on student math performance, the results must be viewed through a critical lens. For example, while machine learning models such as Random Forest and Logistic Regression can reveal significant patterns between features like parental education and household income, the models themselves do not offer causal explanations. Causality can only be inferred through careful design and analysis that controls for potential confounding variables, something that is often difficult to fully address in observational data [27].

Additionally, the interpretation of results must take into account the ethical implications of highlighting certain demographic factors as key predictors of performance. For instance, emphasizing the impact of socio-economic status on performance could lead to oversimplified interpretations that overlook other factors like school quality or individual learning strategies. Therefore, the findings of this study must be communicated with caution, ensuring that the results are presented as potential correlations rather than definitive conclusions about causality.

The study's results demonstrate that socio-economic factors play a meaningful role in predicting students' mathematics performance, with features such as parental education and household income emerging as important predictors. However, the observed differences between male and female students suggest that gender-specific models may capture nuanced patterns not evident in a combined dataset. This finding supports prior literature indicating that learning dynamics can be context-specific and influenced by intersecting socio-economic and demographic factors.

Furthermore, educational stakeholders, including policymakers, curriculum developers, and educators, must carefully consider how the findings are applied to avoid reinforcing stereotypes or contributing to discriminatory practices. The results should be used to inform interventions that target inequalities in education, such as providing additional resources for students from disadvantaged backgrounds, rather than simply documenting disparities without offering solutions [28].

In conclusion, this study highlights the potential of machine learning models to uncover meaningful relationships between socio-economic factors and math performance. However, these models must be used with caution, ensuring that findings inform equitable

and inclusive educational policies. Future research should explore causal mechanisms through longitudinal studies, integrate additional contextual features such as school environment and teacher quality, and investigate intervention strategies that can mitigate the impact of socio-economic disparities on student achievement. Through such an approach, data-driven insights can meaningfully contribute to reducing educational inequalities and promoting better learning outcomes for all students.

Finally, as AI models become more embedded in educational research, questions of accountability arise. Who is responsible when AI-driven decisions lead to harmful outcomes, such as misclassifying students or reinforcing biases in educational assessments? As AI is integrated into education systems, the responsibility for the decisions made by AI models must be clearly defined. In line with emerging AI laws and guidelines, there is an increasing call for transparency and accountability in AI-driven decision-making processes [29].

For this study, accountability extends to ensuring that the data used is accurate, unbiased, and representative of diverse student populations. Moreover, it is essential that the methods used in this research are clearly communicated and accessible, so that stakeholders can trust the findings and use them appropriately. In the context of educational assessment and policy, this also means that decisions based on AI predictions—such as adjusting curriculum or assessment methods—must be subject to human oversight. AI can provide valuable insights into student performance, but it should complement, not replace, human judgment in educational decision-making processes.

References

[1] Wang, M.-T., & Eccles, J. S. (2012). Adolescent behavioral, emotional, and cognitive engagement trajectories in school and their differential relations to educational success.

[2] Enrique Delahoz-Dominguez, Rohemi Zuluaga, Tomas Fontalvo-Herrera, Dataset of academic performance evolution for engineering students, Data in Brief, Volume 30, 2020, 105537, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2020.105537.

[3] Monseur, C., & Lafontaine, D. (2012). Impact of educational systems on the mathematics achievement of boys and girls in Europe: TIMSS and PISA surveys. British Educational Research Journal, 38(6), 989–1017.

[4] Schukajlow, S., Kolter, J., & Blum, W. (2018). Effects of authentic contexts and multiple solution methods on students' performance and motivation in mathematics. Educational Studies in Mathematics, 97(2), 119–137.

[5] Heinze, A., & Reiss, K. (2007). Tasks in mathematics classrooms and their impact on learning. ZDM Mathematics Education, 39(5–6), 371–381.

[6] Prediger, S., Wilhelm, A., Büchter, A., & Gürsoy, E. (2019). Aligning assessments and instruction: The role of test tasks in mathematics education. Journal for Educational Research Online, 11(1), 73–96.

[7] Stankov, L., Morony, S., & Lee, Y. P. (2018). Confidence: The missing link between self-concept and achievement? Educational Psychology, 38(6), 688–707.

[8] Judith M. Delaney, Paul J. Devereux, Math matters! The importance of mathematical and verbal skills for degree performance, Economics Letters, Volume 186, 2020, 108850, ISSN 0165-1765, https://doi.org/10.1016/j.econlet.2019.108850.

[9] Jorge Chica-Olmo, Angeles Sánchez, Fabio H. Sepúlveda-Murillo, Assessing Colombia's policy of socio-economic stratification: An intra-city study of self-reported quality of life, Cities, Volume 97, 2020, 102560, ISSN 0264-2751, https://doi.org/10.1016/j.cities.2019.102560.

[10] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

[11] Wan Yit Yong, Rajesh Jaiswal, and Fernando Perez Tellez. 2023. Explainability in NLP model: Detection of Covid-19 Twitter Fake News. In Proceedings of the 2023 Conference on Human Centered Artificial Intelligence: Education and Practice (HCAIep '23). Association for Computing Machinery, New York, NY, USA, 1. https://doi.org/10.1145/3633083.3633212

[12] Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, Lee MJ, Asadi H. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. AJR Am J Roentgenol. 2019 Jan;212(1):38-43. doi: 10.2214/AJR.18.20224. Epub 2018 Oct 17. Erratum in: AJR Am J Roentgenol. 2019 Feb;212(2):479. doi: 10.2214/AJR.18.20994. PMID: 30332290.

[13] Rainio, O., Teuho, J. & Klén, R. Evaluation metrics and statistical tests for machine learning. Sci Rep 14, 6086 (2024). https://doi.org/10.1038/s41598-024-56706-x

[14] Byun, S. yong, Irvin, M. J., & Bell, B. A. (2014). Advanced Math Course Taking: Effects on Math Achievement and College Enrollment. The Journal of Experimental Education, 83(4), 439–468. https://doi.org/10.1080/00220973.2014.919570

[15] Brahim, G.B. Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features. Arab J Sci Eng 47, 10225–10243 (2022). https://doi.org/10.1007/s13369-021-06548-w

[16] https://6sense.com/tech/learning-management-systems/d2l-brightspace-market-share

[17] https://www.businessofapps.com/data/chegg-statistics

[18] Shokoohi, Roya & Hanif, Noor & Md Dali, Melasutra. (2012). Influence of the Socio-Economic Factors on Children's School Travel. Procedia - Social and Behavioral Sciences. 50. 135–147. 10.1016/j.sbspro.2012.08.022.

[19] Živković, M., Pellizzoni, S., Doz, E. et al. Math self-efficacy or anxiety? The role of emotional and motivational contribution in math performance. Soc Psychol Educ 26, 579–601 (2023). https://doi.org/10.1007/s11218-023-09760-8

[20] Caviola, S., Toffalini, E., Giofrè, D. et al. Math Performance and Academic Anxiety Forms, from Sociodemographic to Cognitive Aspects: a Meta-analysis on 906,311 Participants. Educ Psychol Rev 34, 363–399 (2022). https://doi.org/10.1007/s10648-021-09618-5

[21] https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20230714-1

[22] Courtney, M., Karakus, M., Ersozlu, Z. et al. The influence of ICT use and related attitudes on students' math and science performance: multilevel analyses of the last decade's PISA surveys. Large-scale Assess Educ 10, 8 (2022). https://doi.org/10.1186/s40536-022-00128-6

[23] Cathy O'Neil. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group, USA.

[24] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. ACM Comput. Surv. 54, 6, Article 115 (July 2022), 35 pages. https://doi.org/10.1145/3457607

[25] European Commission. (2016). General Data Protection Regulation (GDPR). Official Journal of the European Union.

[26] Kuner, Christopher and Bygrave, Lee A. and Docksey, Christopher and Docksey, Christopher and Drechsler, Laura and Tosoni, Luca, The EU General Data Protection Regulation: A Commentary/Update of Selected Articles (May 4, 2021). Available at SSRN: https://ssrn.com/abstract=3839645 or http://dx.doi.org/10.2139/ssrn.3839645

[27] Gelman, A. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge university press.

[28] du Boulay, B. (2022). Artificial Intelligence in Education and Ethics. In: Handbook of Open, Distance and Digital Education. Springer, Singapore. https://doi.org/10.1007/978-981-19-0351-9_6-2

[29] Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

[30] Tatachar, A. V. (2021). Comparative assessment of regression models based on model evaluation metrics. International Journal of Innovative Technology and Exploring Engineering, 8(9), 853-860.

[31] Daniel Asante Otchere, Tarek Omar Arbi Ganat, Jude Oghenerurie Ojero, Bennet Nii Tackie-Otoo, Mohamed Yassir Taki, Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions, Journal of Petroleum Science and Engineering, Volume 208, Part E, 2022, 109244, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2021.109244.

[32] https://360digitmg.com/blog/comparison-gridsearchcv-and-randomsearchcv

[33] Huseyin Ahmetoglu, Resul Das, A comprehensive review on detection of cyber-attacks: Data sets, methods, challenges, and future research directions, Internet of Things, Volume 20, 2022, 100615, ISSN 2542-6605, https://doi.org/10.1016/j.iot.2022.100615.

[34] Creswell, J. W., & Creswell, J. D. (2018). Research design (5th ed.). SAGE Publications.

[35] Rindermann, H., & Baumeister, A. E. E. (2015). Parents' SES vs. parental educational behavior and children's development: A reanalysis of the Hart and Risley study. Learning and Individual Differences, 37, 133–138. https://doi.org/10.1016/j.lindif.2014.12.005